

On the Capacity of the Two-Hop Half-Duplex Relay Channel

Nikola Zlatanov[†], Vahid Jamali[‡], and Robert Schober[‡]

[†] University of British Columbia, Vancouver, Canada, and [‡] Friedrich-Alexander-University Erlangen-Nürnberg, Germany

Abstract—Although extensively investigated, the capacity of the two-hop half-duplex (HD) relay channel is not fully understood. In particular, a capacity expression which can be evaluated straightforwardly is not available and an explicit coding scheme which achieves the capacity is not known either. In this paper, we derive a new expression for the capacity of the two-hop HD relay channel based on a simplified converse. Compared to previous results, this capacity expression can be easily evaluated. Moreover, we propose an explicit coding scheme which achieves the capacity. To achieve the capacity, the relay does not only send information to the destination by transmitting information-carrying symbols but also with the zero symbols resulting from the relay's silence during reception. As examples, we compute the capacities of the two-hop HD relay channel for the cases when the source-relay and relay-destination links are both binary-symmetric channels (BSCs) and additive white Gaussian noise (AWGN) channels, respectively, and numerically compare the capacities with the rates achieved by conventional relaying where the relay receives and transmits in a codeword-by-codeword fashion and switches between reception and transmission in a strictly alternating manner. Our numerical results show that the capacities of the two-hop HD relay channel for BSC and AWGN links are significantly larger than the rates achieved with conventional relaying.

I. INTRODUCTION

The two-hop relay channel is comprised of a source, a relay, and a destination, where the direct link between source and destination is not available. In this channel, the message from the source is first transmitted to the relay, which then forwards it to the destination. Generally, a relay can employ two different modes of reception and transmission, i.e., the full-duplex (FD) mode and the half-duplex (HD) mode. Given the limitations of current radio implementations, ideal FD relaying is not possible due to self-interference. The capacity of the two-hop FD relay channel without self-interference has been derived in [1]. On the other hand, although extensively investigated, the capacity of the two-hop HD relay channel is not fully understood. The reason for this is that a capacity expression which can be straightforwardly evaluated is not available and an explicit coding scheme which achieves the capacity is not known either. Currently, for HD relaying, explicit coding schemes exist only for rates which are strictly smaller than the capacity, see [2] and [3]. To achieve the rates given in [2] and [3], the HD relay receives a codeword in one time slot, decodes the received codeword, and re-encodes and re-transmits the decoded information in the following time slot. However, such fixed switching between reception and transmission at the relay was shown to be suboptimal in [4]. In particular, in [4], it was shown that if the fixed scheduling of reception and transmission at the HD relay is abandoned, then additional information can be encoded in the relay's reception and transmission switching pattern yielding an increase in data rate. In addition, it was shown in [4] that the HD relay channel can be analyzed using the framework developed for the FD relay channel in [1]. In particular, results derived for the FD relay channel in [1] can be directly applied to the HD relay channel. Thereby, using the

converse for the degraded relay channel in [1], the capacity of the two-hop HD relay channel is obtained as [4], [5]

$$C = \max_{p(x_1, x_2)} \min \{I(X_1; Y_1 | X_2), I(X_2; Y_2)\}, \quad (1)$$

where X_1 and X_2 are the inputs at source and relay, respectively, Y_1 and Y_2 are the outputs at relay and destination, respectively, and $p(x_1, x_2)$ is the joint probability mass function (PMF) of X_1 and X_2 . Moreover, it was shown in [4] and [5] that X_2 can be represented as $X_2 = [X'_2, U]$, where U is an auxiliary random variable with two outcomes t and r corresponding to the HD relay transmitting and receiving, respectively. Thereby, (1) can be written equivalently as [4], [6]

$$C = \max_{p(x_1, x'_2, u)} \min \{I(X_1; Y_1 | X'_2, U), I(X'_2, U; Y_2)\}, \quad (2)$$

where $p(x_1, x'_2, u)$ is the joint PMF of X_1 , X'_2 , and U . However, the capacity expressions in (1) and (2), respectively, cannot be evaluated since it is not known how X_1 and X_2 nor X_1 , X'_2 , and U are mutually dependent, i.e., $p(x_1, x_2)$ and $p(x_1, x'_2, u)$ are not known. In fact, the authors of [6, page 2552] state that: “*Despite knowing the capacity expression (i.e., expression (2)), its actual evaluation is elusive as it is not clear what the optimal input distribution $p(x_1, x'_2, u)$ is.*” On the other hand, for the coding scheme that would achieve (1) and (2) if $p(x_1, x_2)$ or $p(x_1, x'_2, u)$ was known, it can be argued that it has to be a decode-and-forward strategy since the two-hop HD relay channel belongs to the class of the degraded relay channels defined in [1]. Thereby, the HD relay should decode any received codewords, map the decoded information to new codewords, and transmit them to the destination. Moreover, it is known from [4] that such a coding scheme requires the HD relay to switch between reception and transmission in a symbol-by-symbol manner, and not in a codeword-by-codeword manner as in [2] and [3]. However, since $p(x_1, x_2)$ and $p(x_1, x'_2, u)$ are not known and since an explicit coding scheme does not exist, it is currently not known how to evaluate (1) and (2) nor how to encode additional information in the relay's reception and transmission switching pattern and thereby achieve (1) and (2).

Motivated by the above discussion, in this paper, we derive a new expression for the capacity of the two-hop HD relay channel based on a simplified converse. In contrast to previous results, this capacity expression can be easily evaluated. Moreover, we propose an explicit coding scheme which achieves the capacity. In particular, we show that achieving the capacity requires the relay indeed to switch between reception and transmission in a symbol-by-symbol manner as predicted in [4]. Thereby, the relay does not only send information to the destination by transmitting information-carrying symbols but also with the zero symbols resulting from the relay's silence during reception. In addition, we propose a modified coding scheme for practical implementation where the HD relay receives and

transmits at the same time as in FD relaying, however, the simultaneous reception and transmission is performed such that self-interference is fully avoided. As examples, we compute the capacities of the two-hop HD relay channel for the cases when the source-relay and relay-destination links are both binary-symmetric channels (BSCs) and additive white Gaussian noise (AWGN) channels, respectively, and we numerically compare the capacities with the rates achieved by conventional relaying where the relay receives and transmits in a codeword-by-codeword fashion and switches between reception and transmission in a strictly alternating manner. Our numerical results show that the capacities of the two-hop HD relay channel for BSC and AWGN links are significantly larger than the rates achieved with conventional relaying.

II. SYSTEM MODEL

The two-hop HD relay channel consists of a source, a HD relay, and a destination, and the direct link between source and destination is not available. Due to the HD constraint, the relay cannot transmit and receive at the same time. In the following, we formally define the channel model.

A. Channel Model

The discrete memoryless two-hop HD relay channel is defined by \mathcal{X}_1 , \mathcal{X}_2 , \mathcal{Y}_1 , \mathcal{Y}_2 , and $p(y_1, y_2 | x_1, x_2)$, where \mathcal{X}_1 and \mathcal{X}_2 are the finite input alphabets at the encoders of the source and the relay, respectively, \mathcal{Y}_1 and \mathcal{Y}_2 are the finite output alphabets at the decoders of the relay and the destination, respectively, and $p(y_1, y_2 | x_1, x_2)$ is the PMF on $\mathcal{Y}_1 \times \mathcal{Y}_2$ for given $x_1 \in \mathcal{X}_1$ and $x_2 \in \mathcal{X}_2$. The channel is memoryless in the sense that given the input symbols for the i -th channel use, the i -th output symbols are independent from all previous input symbols. As a result, the conditional PMF $p(y_1^n, y_2^n | x_1^n, x_2^n)$, where the notation a^n is used to denote the ordered sequence $a^n = (a_1, a_2, \dots, a_n)$, can be factorized as $p(y_1^n, y_2^n | x_1^n, x_2^n) = \prod_{i=1}^n p(y_{1i}, y_{2i} | x_{1i}, x_{2i})$.

For the considered channel and the i -th channel use, let X_{1i} and X_{2i} denote the random variables (RVs) which model the input at source and relay, respectively, and let Y_{1i} and Y_{2i} denote the RVs which model the output at the relay and destination, respectively.

In the following, we model the HD constraint of the relay and discuss its effect on some important PMFs that will be used throughout this paper.

B. Mathematical Modelling of the HD Constraint

Due to the HD constraint of the relay, the input and output symbols of the relay cannot take non-zero values at the same time. More precisely, for each channel use, if the input symbol of the relay is non-zero then the output symbol has to be zero, and vice versa, if the output symbol of the relay is non-zero then the input symbol has to be zero. Hence, the following holds

$$Y_{1i} = \begin{cases} Y'_{1i} & \text{if } X_{2i} = 0 \\ 0 & \text{if } X_{2i} \neq 0. \end{cases} \quad (3)$$

where Y'_{1i} is an RV that take values from the set \mathcal{Y}_1 .

In order to model the HD constraint of the relay more conveniently, we represent the input set of the relay \mathcal{X}_2 as the union of two sets $\mathcal{X}_2 = \mathcal{X}_{2R} \cup \mathcal{X}_{2T}$, where \mathcal{X}_{2R} contains only one element, the zero symbol, and \mathcal{X}_{2T} contains all symbols

in \mathcal{X}_2 except the zero symbol. Note that, because of the HD constraint, \mathcal{X}_2 has to contain the zero symbol. Furthermore, we introduce an auxiliary random variable, denoted by U_i , which takes values from the set $\{t, r\}$, where t and r correspond to the relay transmitting a non-zero symbol and a zero symbol, respectively. Hence, U_i is defined as

$$U_i = \begin{cases} r & \text{if } X_{2i} = 0 \\ t & \text{if } X_{2i} \neq 0. \end{cases} \quad (4)$$

Let us denote the probabilities of the relay transmitting a non-zero and a zero symbol for the i -th channel use as $\Pr\{U_i = t\} = \Pr\{X_{2i} \neq 0\} = P_{U_i}$ and $\Pr\{U_i = r\} = \Pr\{X_{2i} = 0\} = 1 - P_{U_i}$, respectively. We now use (4) and represent X_{2i} as a function of the outcome of U_i . Hence, we have

$$X_{2i} = \begin{cases} 0 & \text{if } U_i = r \\ V_i & \text{if } U_i = t, \end{cases} \quad (5)$$

where V_i is an RV with distribution $p_{V_i}(x_{2i})$ that takes values from the set \mathcal{X}_{2T} , or equivalently, an RV which takes values from the set \mathcal{X}_2 , but with $p_{V_i}(x_{2i} = 0) = 0$. From (5), we obtain

$$p(x_{2i} | U_i = r) = \delta(x_{2i}), \quad (6)$$

$$p(x_{2i} | U_i = t) = p_{V_i}(x_{2i}), \quad (7)$$

where $\delta(x) = 1$ if $x = 0$ and $\delta(x) = 0$ if $x \neq 0$. Furthermore, for the derivation of the capacity, we will also need the conditional PMF $p(x_{1i} | x_{2i} = 0)$ which is the input distribution at the source when relay transmits a zero (i.e., when $U_i = r$) and $p(x_{2i} | U_i = t) = p_{V_i}(x_{2i})$ which is the input distribution at the relay when the relay transmits non-zero symbols. As we will see in Theorem 1, these distributions have to be optimized in order to achieve the capacity. Using $p(x_{2i} | U_i = r)$ and $p(x_{2i} | U_i = t)$, and the law of total probability, the PMF of X_{2i} , $p(x_{2i})$, is obtained as

$$\begin{aligned} p(x_{2i}) &= p(x_{2i} | U_i = t)P_{U_i} + p(x_{2i} | U_i = r)(1 - P_{U_i}) \\ &\stackrel{(a)}{=} p_{V_i}(x_{2i})P_{U_i} + \delta(x_{2i})(1 - P_{U_i}), \end{aligned} \quad (8)$$

where (a) follows from (6) and (7). In addition, we will also need the distribution of Y_{2i} , $p(y_{2i})$, which, using the law of total probability, can be written as

$$p(y_{2i}) = p(y_{2i} | U_i = t)P_{U_i} + p(y_{2i} | U_i = r)(1 - P_{U_i}). \quad (9)$$

On the other hand, using X_{2i} and the law of total probability, $p(y_{2i} | U_i = r)$ can be written as

$$\begin{aligned} p(y_{2i} | U_i = r) &= \sum_{x_{2i} \in \mathcal{X}_2} p(y_{2i}, x_{2i} | U_i = r) \\ &= \sum_{x_{2i} \in \mathcal{X}_2} p(y_{2i} | x_{2i}, U_i = r) p(x_{2i} | U_i = r) \\ &\stackrel{(a)}{=} \sum_{x_{2i} \in \mathcal{X}_2} p(y_{2i} | x_{2i}, U_i = r) \delta(x_{2i}) = p(y_{2i} | x_{2i} = 0, U_i = r) \\ &\stackrel{(b)}{=} p(y_{2i} | x_{2i} = 0), \end{aligned} \quad (10)$$

where (a) is due to (6) and (b) is the result of conditioning on the same variable twice since if $X_{2i} = 0$ then $U_i = r$, and vice versa. On the other hand, using X_{2i} and the law of total

probability, $p(y_{2i}|U_i = t)$ can be written as

$$\begin{aligned} p(y_{2i}|U_i = t) &= \sum_{x_{2i} \in \mathcal{X}_2} p(y_{2i}, x_{2i}|U_i = t) \\ &= \sum_{x_{2i} \in \mathcal{X}_2} p(y_{2i}|x_{2i}, U_i = t)p(x_{2i}|U_i = t) \\ &\stackrel{(a)}{=} \sum_{x_{2i} \in \mathcal{X}_{2T}} p(y_{2i}|x_{2i})p_{V_i}(x_{2i}), \end{aligned} \quad (11)$$

where (a) follows for (7) and since V_i takes values from set \mathcal{X}_{2T} . In (11), $p(y_{2i}|x_{2i})$ is the distribution at the output of the relay-destination channel conditioned on the relay transmitting the symbol x_{2i} .

C. Mutual Information and Entropy

For the capacity expression given later in Theorem 1, we need $I(X_1; Y_1|X_2 = 0)$, which is the mutual information between the source's input X_1 and the relay's output Y_1 conditioned on the relay having its input set to $X_2 = 0$, and $I(X_2; Y_2)$, which is the mutual information between the relay's input X_2 and the destination's output Y_2 .

The mutual information $I(X_1; Y_1|X_2 = 0)$ is obtained by definition as

$$\begin{aligned} I(X_1; Y_1|X_2 = 0) &= \sum_{x_1 \in \mathcal{X}_1} \sum_{y_1 \in \mathcal{Y}_1} p(y_1|x_1, x_2 = 0) \\ &\times p(x_1|x_2 = 0) \log_2 \left(\frac{p(y_1|x_1, x_2 = 0)}{p(y_1|x_2 = 0)} \right), \end{aligned} \quad (12)$$

where

$$p(y_1|x_2 = 0) = \sum_{x_1 \in \mathcal{X}_1} p(y_1|x_1, x_2 = 0)p(x_1|x_2 = 0). \quad (13)$$

In (12) and (13), $p(y_1|x_1, x_2 = 0)$ is the distribution at the output of the source-relay channel conditioned on the relay having its input set to $X_2 = 0$, and conditioned on the input symbols at the source X_1 .

On the other hand, in order to obtain $I(X_2; Y_2)$, we use the following identity

$$I(X_2; Y_2) = H(Y_2) - H(Y_2|X_2), \quad (14)$$

where $H(Y_2)$ is the entropy of RV Y_2 , and $H(Y_2|X_2)$ is the entropy of Y_2 conditioned on the knowledge of X_2 . The entropy $H(Y_2)$ can be found by definition as

$$\begin{aligned} H(Y_2) &= - \sum_{y_2 \in \mathcal{Y}_2} p(y_2) \log_2(p(y_2)) \\ &\stackrel{(a)}{=} - \sum_{y_2 \in \mathcal{Y}_2} [p(y_2|U = t)P_U + p(y_2|U = r)(1 - P_U)] \\ &\times \log_2 [p(y_2|U = t)P_U + p(y_2|U = r)(1 - P_U)], \end{aligned} \quad (15)$$

where (a) follows from (9). Now, inserting $p(y_2|U = r)$ and $p(y_2|U = t)$ given in (10) and (11), respectively, into (15), we obtain the final expression for $H(Y_2)$, as

$$\begin{aligned} H(Y_2) &= - \sum_{y_2 \in \mathcal{Y}_2} \left[P_U \sum_{x_2 \in \mathcal{X}_{2T}} p(y_2|x_2)p_V(x_2) \right. \\ &\quad \left. + p(y_2|x_2 = 0)(1 - P_U) \right] \\ &\times \log_2 \left[P_U \sum_{x_2 \in \mathcal{X}_{2T}} p(y_2|x_2)p_V(x_2) + p(y_2|x_2 = 0)(1 - P_U) \right]. \end{aligned} \quad (16)$$

On the other hand, the conditional entropy $H(Y_2|X_2)$ can be found based on its definition as

$$\begin{aligned} H(Y_2|X_2) &= - \sum_{x_2 \in \mathcal{X}_2} p(x_2) \sum_{y_2 \in \mathcal{Y}_2} p(y_2|x_2) \log_2(p(y_2|x_2)) \\ &\stackrel{(a)}{=} - P_U \sum_{x_2 \in \mathcal{X}_{2T}} p_V(x_2) \sum_{y_2 \in \mathcal{Y}_2} p(y_2|x_2) \log_2(p(y_2|x_2)) \\ &\quad - (1 - P_U) \sum_{y_2 \in \mathcal{Y}_2} p(y_2|x_2 = 0) \log_2(p(y_2|x_2 = 0)), \end{aligned} \quad (17)$$

where (a) follows by inserting $p(x_2)$ given in (8). Inserting $H(Y_2)$ and $H(Y_2|X_2)$ given in (16) and (17), respectively, into (14), we obtain the final expression for $I(X_2; Y_2)$, which is dependent on $p(x_2)$, i.e., on $p_V(x_2)$ and P_U . To emphasize the dependance of $I(X_2; Y_2)$ on P_U , we sometimes write $I(X_2; Y_2)$ as $I(X_2; Y_2)|_{P_U}$.

We are now ready to present the capacity of the considered channel.

III. CAPACITY

In this section, we provide an easy-to-evaluate expression for the capacity of the two-hop HD relay channel, an explicit coding scheme that achieves the capacity, and the converse for the capacity.

A. The Capacity

A new expression for the capacity of the two-hop HD relay channel is given in the following theorem.

Theorem 1: The capacity of the two-hop HD relay channel is given by

$$\begin{aligned} C &= \max_{P_U} \min \left\{ \max_{p(x_1|x_2=0)} I(X_1; Y_1|X_2 = 0)(1 - P_U), \right. \\ &\quad \left. \max_{p_V(x_2)} I(X_2; Y_2)|_{P_U} \right\}, \end{aligned} \quad (18)$$

where $I(X_1; Y_1|X_2 = 0)$ is given in (12) and $I(X_2; Y_2)$ is given in (14)-(17). The optimal P_U that maximizes the capacity in (18) is given by $P_U^* = \min\{P_U', P_U''\}$, where P_U' and P_U'' are the solutions of

$$\max_{p(x_1|x_2=0)} I(X_1; Y_1|X_2 = 0)(1 - P_U) = \max_{p_V(x_2)} I(X_2; Y_2)|_{P_U} \quad (19)$$

and

$$\frac{\partial \left(\max_{p_V(x_2)} I(X_2; Y_2)|_{P_U} \right)}{\partial P_U} = 0, \quad (20)$$

respectively. If $P_U^* = P_U'$, then both terms inside the $\min\{\cdot\}$ function of the capacity in (18) become identical. Whereas, if $P_U^* = P_U''$, then the capacity in (18) simplifies to

$$C = \max_{p_V(x_2)} I(X_2; Y_2)|_{P_U=P_U''} = \max_{p(x_2)} I(X_2; Y_2), \quad (21)$$

which is the capacity of the relay-destination channel.

Proof: The proof of the capacity given in (18) is provided in two parts. In the first part, given in Section III-B, we show that there exists a coding scheme that achieves a rate R which is smaller, but arbitrarily close to capacity C . In the second part, given in Section III-C, we prove that any rate R for which the probability of error can be made arbitrarily small, must be smaller than capacity C given in (18). The rest of the theorem follows from solving (18) with respect to P_U ,

and simplifying the result. In particular, note that the first term inside the $\min\{\cdot\}$ function in (18) is a decreasing function with respect to P_U . This function achieves its maximum for $P_U = 0$ and its minimum, which is zero, for $P_U = 1$. On the other hand, the second term inside the $\min\{\cdot\}$ function in (18) is a concave function with respect to P_U . To see this, note that $I(X_2; Y_2)$ is a concave function with respect to $p(x_2)$, i.e., with respect to the vector comprised of the probabilities in $p(x_2)$, see [7]. Now, since $1 - P_U$ is just the probability $p(x_2 = 0)$ and since $p_V(x_2)$ contains the rest of the probability constrained parameters in $p(x_2)$, $I(X_2; Y_2)$ is a jointly concave function with respect to $p_V(x_2)$ and P_U . In [8, pp. 87-88], it is proven that if $f(x, y)$ is a jointly concave function in both (x, y) and \mathcal{C} is a convex nonempty set, then the function $g(x) = \max_{y \in \mathcal{C}} f(x, y)$ is concave in x . Using this result, and noting that the domain of $p_V(x_2)$ is specified by the probability constraints, i.e., by a convex nonempty set, we can directly conclude that $\max_{p_V(x_2)} I(X_2; Y_2)|_{P_U}$ is concave with respect to P_U .

Now, the maximization of the minimum of the decreasing and concave functions with respect to P_U , given in (18), has a solution $P_U = P_U''$, when the concave function reaches its maximum, found from (20), and when for this point, i.e., for $P_U = P_U''$, the decreasing function is larger than the concave function. Otherwise, the solution is $P_U = P_U'$ which is the point when the decreasing and concave functions intersect, which is found from (19) and in which case $P_U' < P_U''$ holds. We note that (19) has only one solution since for $P_U = 1$ the left term in (19) becomes zero. Whereas, for $P_U = 0$, $\max_{p_V(x_2)} I(X_2; Y_2)|_{P_U=0} = 0$, since $p(x_2 = 0) = 1$ occurs, and for $P_U = 1$, $\max_{p_V(x_2)} I(X_2; Y_2)|_{P_U=1} \geq 0$, where equality holds if and only if for $P_U = 1$, $p_V(x_2)$ becomes a degenerate (or deterministic) PMF. ■

B. Achievability of the Capacity

In the following, we describe a method for transferring nR bits of information in $n + k$ channel uses, where $n, k \rightarrow \infty$ and $n/(n + k) \rightarrow 1$ as $n, k \rightarrow \infty$. As a result, the information is transferred at rate R . To this end, the transmission is carried out in $N + 1$ blocks, where $N \rightarrow \infty$. In each block, we use the channel k times. The numbers N and k are chosen such that $n = Nk$ holds.

We transmit message W , drawn uniformly from message set $\{1, 2, \dots, 2^{nR}\}$, from the source via the HD relay to the destination. To this end, before the start of transmission, message W is split into N messages, denoted by $w(1), \dots, w(N)$, where each $w(i)$, $\forall i$, contains kR bits. The transmission is carried out in the following manner. In block one, the source sends message $w(1)$ in k channel uses to the relay and the relay is silent. In block i , for $i = 2, \dots, N$, source and relay send messages $w(i)$ and $w(i - 1)$ to relay and destination, respectively, in k channel uses. In block $N + 1$, the relay sends message $w(N)$ in k channel uses to the destination and the source is silent. Hence, in the first block and in the $(N + 1)$ -th block the relay and the source are silent, respectively, since in the first block the relay does not have information to transmit, and in block $N + 1$, the source has no more information to transmit. In blocks 2 to N , both source and relay transmit, while meeting the HD constraint in every channel use. Hence, during the $N + 1$

blocks, the channel is used $k(N + 1)$ times to send $nR = NkR$ bits of information, leading to an overall information rate given by $\lim_{N \rightarrow \infty} \lim_{k \rightarrow \infty} \frac{NkR}{k(N + 1)} = R$ bits/use.

A detailed description of the proposed coding scheme is given in the following, where we explain the rates, codebooks, encoding, and decoding used for transmission.

Rates: The transmission rate of both source and relay is denoted by R and given by

$$R = C - \epsilon, \quad (22)$$

where C is given in Theorem 1 and $\epsilon > 0$ is an arbitrarily small number. Note that R is a function of P_U^* , see Theorem 1.

Codebooks: We have two codebooks: The source's transmission codebook and the relay's transmission codebook.

The source's transmission codebook is generated by mapping each possible binary sequence comprised of kR bits, where R is given by (22), to a codeword¹ $\mathbf{x}_{1|r}$ comprised of $k(1 - P_U^*)$ symbols. The symbols in each codeword $\mathbf{x}_{1|r}$ are generated independently according to distribution $p(x_1|x_2 = 0)$. Since in total there are 2^{kR} possible binary sequences comprised of kR bits, with this mapping we generate 2^{kR} codewords $\mathbf{x}_{1|r}$ each containing $k(1 - P_U^*)$ symbols. These 2^{kR} codewords form the source's transmission codebook, which we denote by $\mathcal{C}_{1|r}$.

The relay's transmission codebook is generated by mapping each possible binary sequence comprised of kR bits, where R is given by (22), to a transmission codeword \mathbf{x}_2 comprised of k symbols. The i -th symbol, $i = 1, \dots, k$, in codeword \mathbf{x}_2 is generated in the following manner. For each symbol a coin is tossed. The coin is such that it produces symbol r with probability $1 - P_U^*$ and symbol t with probability P_U^* . If the outcome of the coin flip is r , then the i -th symbol of the relay's transmission codeword \mathbf{x}_2 is set to zero. Otherwise, if the outcome of the coin flip is t , then the i -th symbol of codeword \mathbf{x}_2 is generated independently according to distribution $p_V(x_2)$. The 2^{kR} codewords \mathbf{x}_2 form the relay's transmission codebook denoted by \mathcal{C}_2 .

The two codebooks are known at all three nodes.

Encoding, Transmission, and Decoding: In the first block, the source maps $w(1)$ to the appropriate codeword $\mathbf{x}_{1|r}(1)$ from its codebook $\mathcal{C}_{1|r}$. Then, codeword $\mathbf{x}_{1|r}(1)$ is transmitted to the relay, which is scheduled to always receive and be silent (i.e., sets its input to zero) during the first block. However, knowing that the transmitted codeword from the source $\mathbf{x}_{1|r}(1)$ is comprised of $k(1 - P_U^*)$ symbols, the relay constructs the received codeword, denoted by $\mathbf{y}_{1|r}(1)$, only from the first $k(1 - P_U^*)$ received symbols. In [9, Appendix A], we prove that codeword $\mathbf{x}_{1|r}(1)$ sent in the first block can be decoded successfully from the received codeword at the relay $\mathbf{y}_{1|r}(1)$ using a typical decoder [7] since R satisfies

$$R < \max_{p(x_1|x_2=0)} I(X_1; Y_1|X_2 = 0)(1 - P_U^*). \quad (23)$$

In blocks $i = 2, \dots, N$, the encoding, transmission, and decoding are performed as follows. In blocks $i = 2, \dots, N$, the source and the relay map $w(i)$ and $w(i - 1)$ to the appropriate codewords $\mathbf{x}_{1|r}(i)$ and $\mathbf{x}_2(i)$ from codebooks $\mathcal{C}_{1|r}$ and \mathcal{C}_2 , respectively. Note that the source also knows $\mathbf{x}_2(i)$ since $\mathbf{x}_2(i)$

¹The subscript $1|r$ in $\mathbf{x}_{1|r}$ is used to indicate that codeword $\mathbf{x}_{1|r}$ is comprised of symbols which are transmitted by the source only when $U_i = r$.

was generated from $w(i-1)$ which the source transmitted in the previous (i.e., $(i-1)$ -th) block. The transmission of $\mathbf{x}_{1|r}(i)$ and $\mathbf{x}_2(i)$ can be performed in two ways: 1) by the relay switching between reception and transmission, and 2) by the relay always receiving and transmitting as in FD relaying. We first explain the first option.

Note that both source and relay know the position of the zero symbols in $\mathbf{x}_2(i)$. Hence, if the first symbol in codeword $\mathbf{x}_2(i)$ is zero, then in the first symbol interval of block i , the source transmits its first symbol from codeword $\mathbf{x}_{1|r}(i)$ and the relay receives. By receiving, the relay actually also sends the first symbol of codeword $\mathbf{x}_2(i)$, which is the symbol zero, i.e., $x_{21} = 0$. On the other hand, if the first symbol in codeword $\mathbf{x}_2(i)$ is non-zero, then in the first symbol interval of block i , the relay transmits its first symbol from codeword $\mathbf{x}_2(i)$ and the source is silent. The same procedure is performed for the j -th channel use in block i , for $j = 1, \dots, k$. In particular, if the j -th symbol in codeword $\mathbf{x}_2(i)$ is zero, then in the j -th channel use of block i the source transmits its next untransmitted symbol from codeword $\mathbf{x}_{1|r}(i)$ and the relay receives. With this reception, the relay actually also sends the j -th symbol of codeword $\mathbf{x}_2(i)$, which is the symbol zero, i.e., $x_{2j} = 0$. On the other hand, if the j -th symbol in codeword $\mathbf{x}_2(i)$ is non-zero, then for the j -th channel use of block i , the relay transmits the j -th symbol of codeword $\mathbf{x}_2(i)$ and the source is silent. Since codeword $\mathbf{x}_2(i)$ contains approximately $k(1 - P_U^*)$ symbols zeros, for $k \rightarrow \infty$ the source can transmit practically all² of its $k(1 - P_U^*)$ symbols from codeword $\mathbf{x}_{1|r}(i)$ during a single block to the relay. Let $\mathbf{y}_{1|r}(i)$ denote the corresponding received codeword at the relay. In [9, Appendix A], we prove that the codewords $\mathbf{x}_{1|r}(i)$ sent in blocks $i = 2, \dots, N$ can be decoded successfully at the relay from the corresponding received codewords $\mathbf{y}_{1|r}(i)$ using a typical decoder [7] since R satisfies (23). On the other hand, the relay sends the entire codeword $\mathbf{x}_2(i)$, comprised of k symbols of which approximately $k(1 - P_U^*)$ are zeros, to the destination. In particular, the relay sends the $k(1 - P_U^*)$ zero symbols of codeword $\mathbf{x}_2(i)$ to the destination by being silent during reception, and sends the remaining kP_U^* symbols of codeword $\mathbf{x}_2(i)$ to the destination by actually transmitting them. On the other hand, the destination listens during the entire block and receives a codeword $\mathbf{y}_2(i)$. In [9, Appendix B], we prove that the destination can successfully decode $\mathbf{x}_2(i)$ from the received codeword $\mathbf{y}_2(i)$, and thereby obtain $w(i-1)$, since rate R satisfies

$$R < \max_{p_V(x_2)} I(X_2; Y_2) \Big|_{P_U = P_U^*}. \quad (24)$$

In a practical implementation, the relay may not be able to switch between reception and transmission in a symbol-by-symbol manner, due to practical constraints regarding the speed of switching. Instead, we may allow the relay to receive and transmit at the same time and in the same frequency band similar to FD relaying. However, this simultaneous reception and transmission is performed while avoiding self-interference

²Due to the strong law of large numbers, the number of zero symbols in $\mathbf{x}_2(i)$ is $k(1 - P_U) - \varepsilon$, where ε is an integer satisfying $\lim_{k \rightarrow \infty} \varepsilon/k = 0$ [7]. Hence, when we say that the source can transmit practically all of its symbols, we mean either all or all except for a negligible fraction $\lim_{k \rightarrow \infty} \varepsilon/k = 0$ of them. This fraction is negligible such that the decisions of the typical decoder are not affected for $k \rightarrow \infty$, see [9, Appendix A].

since, in each symbol interval, either the input or the output information-carrying symbol of the relay is zero. This is accomplished in the following manner. The source performs the same operations as for the case when the relay switches between reception and transmission. On the other hand, the relay transmits all symbols from $\mathbf{x}_2(i)$ while continuously listening. Then, the relay discards from the received codeword, denoted by $\mathbf{y}_1(i)$, those symbols for which the corresponding symbols in $\mathbf{x}_2(i)$ are non-zero, and only collects the symbols in $\mathbf{y}_1(i)$ for which the corresponding symbols in $\mathbf{x}_2(i)$ are equal to zero. From the collected symbols in $\mathbf{y}_1(i)$, the relay obtains the received information-carrying codeword $\mathbf{y}_{1|r}(i)$ which it needs for decoding. Codeword $\mathbf{y}_{1|r}(i)$ is completely free of self-interference since the symbols in $\mathbf{y}_{1|r}(i)$ were received in symbol intervals for which the corresponding transmit symbol at the relay was zero.

In the last (i.e., the $(N+1)$ -th) block, the source is silent and the relay transmits $w(N)$ by mapping it to the corresponding codeword $\mathbf{x}_2(i)$ from set \mathcal{C}_2 . The relay transmits all symbols in codeword $\mathbf{x}_2(i)$ to the destination. The destination can decode the received codeword in block $N+1$ successfully, since (24) holds.

Finally, since both relay and destination can decode their respective codewords in each block, the entire message W can be decoded successfully at the destination at the end of the $(N+1)$ -th block.

C. Converse

As shown in [4], the HD relay channel can be investigated with the framework developed for the FD relay channel in [1]. Since the considered two-hop HD relay channel belongs to the class of degraded relay channels defined in [1], the rate of this channel is upper bounded by [1], [4]

$$R \leq \max_{p(x_1, x_2)} \min \{I(X_1; Y_1|X_2), I(X_2; Y_2)\}. \quad (25)$$

On the other hand, $I(X_1; Y_1|X_2)$ can be simplified as

$$\begin{aligned} I(X_1; Y_1|X_2) &= I(X_1; Y_1|X_2 = 0)(1 - P_U) + I(X_1; Y_1|X_2 \neq 0)P_U \\ &\stackrel{(a)}{=} I(X_1; Y_1|X_2 = 0)(1 - P_U), \end{aligned} \quad (26)$$

where (a) follows from (3) since when $X_2 \neq 0$, Y_1 is deterministically zero thereby leading to $I(X_1; Y_1|X_2 \neq 0) = 0$. Inserting (26) into (25), (25) simplifies as

$$R \leq \max_{p(x_1, x_2)} \min \{I(X_1; Y_1|X_2 = 0)(1 - P_U), I(X_2; Y_2)\}. \quad (27)$$

Since $p(x_1, x_2) = p(x_1|x_2)p(x_2)$, where $p(x_2)$ is given in (8) as a function of P_U and $p_V(x_2)$, the maximization in (27) with respect to $p(x_1, x_2)$ can be resolved into joint maximization with respect to $p(x_1|x_2)$, $p_V(x_2)$, and P_U . Now, since $I(X_1; Y_1|X_2 = 0)(1 - P_U)$ and $I(X_2; Y_2)$ are functions of $p(x_1|x_2 = 0)$ and $p_V(x_2)$, respectively, and no other function inside the $\min\{\cdot\}$ function in (27) is dependent on the distributions $p(x_1|x_2 = 0)$ and $p_V(x_2)$, (27) can be written equivalently as

$$R \leq \max_{P_U} \min \left\{ \max_{p(x_1|x_2=0)} I(X_1; Y_1|X_2 = 0)(1 - P_U), \max_{p_V(x_2)} I(X_2; Y_2) \right\}, \quad (28)$$

where $\max_{p(x_1|x_2=0)} I(X_1; Y_1|X_2=0)$ and $\max_{p_V(x_2)} I(X_2; Y_2)$ exist since these functions are concave with respect to $p(x_1|x_2=0)$ and $p_V(x_2)$, respectively. On the other hand, the maximum in (28) with respect to P_U exists since the first and the second terms inside the $\min\{\cdot\}$ function in (28) are monotonically decreasing and concave functions with respect to P_U , respectively (see proof of Theorem 1 for concavity). This concludes the proof of the converse.

IV. APPLICATION EXAMPLES: BSC AND AWGN

In this section, we use Theorem 1 to derive the capacity of the two-hop HD relay channel for the cases when the source-relay and relay-destination links are both BSCs and AWGN channels, respectively.

A. BSC

Assume that the source-relay and relay-destination links are both BSCs with probability of error $P_{\varepsilon 1}$ and $P_{\varepsilon 2}$, respectively. Let $H(P) = -P \log_2(P) - (1-P) \log_2(1-P)$ denote the binary entropy function. Then, the capacity for this channel is given in the following corollary.

Corollary 1: The capacity of the considered relay channel with BSCs links is given by

$$C = \max_{P_U} \min\{(1 - H(P_{\varepsilon 1}))(1 - P_U), -A \log_2(A) - (1 - A) \log_2(1 - A) - H(P_{\varepsilon 2})\}, \quad (29)$$

where $A = P_{\varepsilon 2}(1 - 2P_U) + P_U$, and is achieved with

$$p_V(x_2) = \delta(x_2 - 1) \quad (30)$$

$$p(x_1 = 0|x_2 = 0) = p(x_1 = 1|x_2 = 0) = 1/2. \quad (31)$$

There are two cases for the optimal P_U which maximizes (29). If P_U found from

$$(1 - H(P_{\varepsilon 1}))(1 - P_U) = -A \log_2(A) - (1 - A) \log_2(1 - A) - H(P_{\varepsilon 2}) \quad (32)$$

is smaller than $1/2$, then the optimal P_U which maximizes (29) is found as the solution to (32). In this case, the first and second term inside the $\min\{\cdot\}$ of the capacity become equal. Otherwise, if P_U found from (32) is $P_U \geq 1/2$, then the optimal P_U which maximizes (29) is $P_U = 1/2$, and the capacity simplifies to

$$C = 1 - H(P_{\varepsilon 2}), \quad (33)$$

which is the capacity of the relay-destination link.

Proof: Please refer to [9, Section IV-A]. ■

B. AWGN

We now assume that the source-relay and relay-destination links are AWGN channels, i.e., channels which are impaired by independent, real-valued, zero-mean AWGN with variances σ_1^2 and σ_2^2 , respectively. More precisely, the outputs at the relay and the destination are given by $Y_k = X_k + N_k$, $k \in \{1, 2\}$, where N_k is a zero-mean Gaussian RV with variance σ_k^2 , $k \in \{1, 2\}$, with distribution $p_{N_k}(z)$, $k \in \{1, 2\}$, $-\infty \leq z \leq \infty$. Moreover, assume that the symbols transmitted by

the source and the relay must satisfy the following average power constraints³

$$\sum_{x_1 \in \mathcal{X}_1} x_1^2 p(x_1|x_2=0) \leq P_1 \quad \text{and} \quad \sum_{x_2 \in \mathcal{X}_{2T}} x_2^2 p_V(x_2) \leq P_2. \quad (34)$$

Then, the capacity for this channel is given in the following corollary.

Corollary 2: The capacity of the considered relay channel where the source-relay and relay-destination links are both AWGN channels with noise variances σ_1^2 and σ_2^2 , respectively, and where the average power constraints of the inputs of source and relay are given by (34), is given by

$$C = \frac{1}{2} \log_2 \left(1 + \frac{P_1}{\sigma_1^2} \right) (1 - P_U^*) \\ \stackrel{(a)}{=} - \int_{-\infty}^{\infty} \left(P_U^* \sum_{k=1}^K p_k^* p_{N_2}(y_2 - x_{2k}^*) + (1 - P_U^*) p_{N_2}(y_2) \right) \\ \times \log_2 \left(P_U^* \sum_{k=1}^K p_k^* p_{N_2}(y_2 - x_{2k}^*) + (1 - P_U^*) p_{N_2}(y_2) \right) dy_2 \\ - \frac{1}{2} \log_2(2\pi e \sigma_2^2), \quad (35)$$

where the optimal P_U^* is found such that equality (a) in (35) holds. The capacity in (35) is achieved when $p(x_1|x_2=0)$ is the zero-mean Gaussian distribution with variance P_1 and $p_V^*(x_2) = \sum_{k=1}^K p_k^* \delta(x_2 - x_{2k}^*)$ is a discrete distribution which satisfies

$$\sum_{k=1}^K p_k^* = 1 \quad \text{and} \quad \sum_{k=1}^K p_k^* (x_{2k}^*)^2 = P_2. \quad (36)$$

and maximizes $H(Y_2)$.

Proof: Please refer to [9, Section IV-B]. ■

Unfortunately, obtaining the optimal $p_V(x_2)$ which satisfies (36) and maximizes $H(Y_2)$ in closed form is difficult, if not impossible, see [9, Section IV-B] for more details. Therefore, a brute-force search has to be used in order to find x_{2k}^* and p_k^* , $\forall k$. Instead of an optimal discrete input distribution at the relay $p_V^*(x_2)$, we can use⁴ a continuous, zero-mean Gaussian distribution with variance P_2 , which will produce a rate smaller than the capacity, given by

$$R_{\text{Gauss}} = \frac{1}{2} \log_2 \left(1 + \frac{P_1}{\sigma_1^2} \right) (1 - P_U) \\ \stackrel{(a)}{=} - \int_{-\infty}^{\infty} (P_U p_G(y_2) + (1 - P_U) p_{N_2}(y_2)) \\ \times \log_2 (P_U p_G(y_2) + (1 - P_U) p_{N_2}(y_2)) dy_2 - \frac{1}{2} \log_2(2\pi e \sigma_2^2), \quad (37)$$

where P_U is found such that equality (a) holds and $p_G(y_2)$ is a continuous, zero-mean Gaussian distribution with variance $P_2 + \sigma_2^2$. In the following, we numerically evaluate the capacities in Corollaries 1 and 2.

³If the optimal distributions $p(x_1|x_2=0)$ and $p_V(x_2)$ turn out to be continuous, the sums in (34) should be replaced by integrals.

⁴Note that when $p_V(x_2)$ is a continuous Gaussian distribution, the probability that $x_2 = 0$ will occur is zero. Hence, the definition of $p_V(x_2)$ is not violated.

V. NUMERICAL EXAMPLES

In this section, we numerically evaluate the capacities of the two-hop HD relay channel when the source-relay and relay-destination links are both BSCs and AWGN channels, respectively, and compare it to the maximal achievable rates of conventional relaying [2], [3].

A. BSC

For simplicity, we assume symmetric links with $P_{\varepsilon 1} = P_{\varepsilon 2} = P_{\varepsilon}$. The capacity is given by (29). This capacity is plotted in Fig. 1, where the optimal P_U is found from (32) using a mathematical software package, e.g. Mathematica. As a benchmark, in Fig. 1, we also show the maximal achievable rate using conventional relaying, obtained as $R_{\text{conv}} = (1 - H(P_{\varepsilon}))/2$. As can be seen from Fig. 1, the capacity is significantly higher than the maximal rate of conventional relaying. For example, when both links are error-free, i.e., $P_{\varepsilon} = 0$, conventional relaying achieves 0.5 bits/channel use, whereas the capacity is 0.77291, which is 54% larger than the rate achieved with conventional relaying. We note that this value was also reported in [5, page 327], where only the case of error-free BSCs was investigated.

B. AWGN

For the AWGN case, the capacity is evaluated based on Corollary 2. However, since for this case the optimal input distribution at the relay $p_V(x_2)$ is unknown, i.e., the values of p_k^* and x_{2k}^* in (35) are unknown, we have performed a brute force search for the values of p_k^* and x_{2k}^* which maximize (35). Two examples of such distributions are shown in [9, Fig. 6] for two different values of the SNR $P_1/\sigma_1^2 = P_2/\sigma_2^2$.

The capacity is shown in Fig. 2 for the case when $P_1/\sigma_1^2 = P_2/\sigma_2^2 = P/\sigma^2$. In Fig. 2, we also show the rate achieved when instead of an optimal discrete input distribution at the relay $p_V^*(x_2)$, we use a continuous, zero-mean Gaussian distribution with variance P_2 , in which case the rate is given in (37). From Fig. 2, we can see that R_{Gauss} is smaller than the capacity, which was expected. However, the loss in performance caused by the Gaussian inputs is moderate, which suggests that the performance gains obtained by the proposed coding scheme are mainly due to the exploitation of the silent (zero) symbols for conveying information from the HD relay to the destination rather than the optimization of $p_V(x_2)$. As benchmark, in Fig. 2, we have also shown the maximal achievable rate using conventional relaying, obtained for $P_1/\sigma_1^2 = P_2/\sigma_2^2 = P/\sigma^2$ as [2], [3] $R_{\text{conv}} = \log_2(1 + P/\sigma^2)/4$. Comparing the capacity with R_{conv} in Fig. 2, we see that for $10 \text{ dB} \leq P/\sigma^2 \leq 30 \text{ dB}$, 3 to 5 dB gain is achieved. Hence, large performance gains are achieved using the proposed capacity coding scheme even if suboptimal input distributions at the relay are employed.

VI. CONCLUSIONS

We have derived an easy-to-evaluate expression for the capacity of the two-hop HD relay channel based on a simplified converse. Moreover, we have proposed an explicit coding scheme which achieves the capacity. In particular, we showed that the capacity is achieved when additional information is sent by the relay to the destination using the zero symbol implicitly sent by the relay's silence during reception. Furthermore, we have evaluated the capacity for the cases when both links are BSCs and AWGN channels, respectively. From

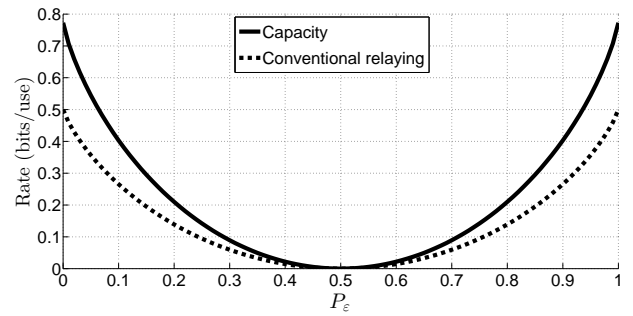


Fig. 1. Comparison of rates for the BSC as a function of the error probability $P_{\varepsilon 1} = P_{\varepsilon 2} = P_{\varepsilon}$.

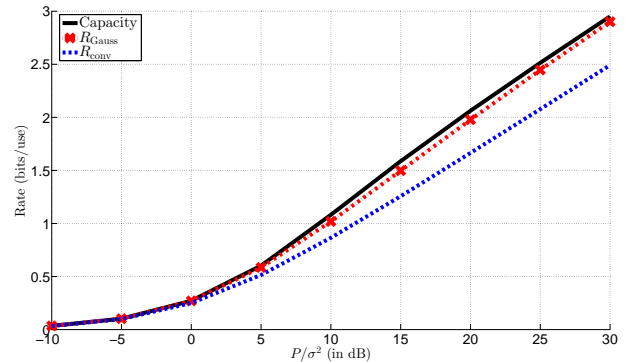


Fig. 2. Source-relay and relay destination links are AWGN channels with $P_1/\sigma_1^2 = P_2/\sigma_2^2 = P/\sigma^2$.

the numerical examples, we have observed that the capacity is significantly larger than the rate achieved with conventional relaying protocols.

REFERENCES

- [1] T. Cover and A. El Gamal, "Capacity Theorems for the Relay Channel," *IEEE Trans. Inform. Theory*, vol. 25, pp. 572–584, Sep. 1979.
- [2] A. Host-Madsen and J. Zhang, "Capacity Bounds and Power Allocation for Wireless Relay Channels," *IEEE Trans. Inform. Theory*, vol. 51, pp. 2020–2040, Jun. 2005.
- [3] M. Khojastepour, A. Sabharwal, and B. Aazhang, "On the Capacity of 'Cheap' Relay Networks," in *Proc. Conf. on Inform. Sciences and Systems*, Princeton, NJ, 2003.
- [4] G. Kramer, "Models and Theory for Relay Channels with Receive Constraints," in *Proc. 42nd Annual Allerton Conf. on Commun., Control, and Computing*, Monticello, IL, 2004, pp. 1312–1321.
- [5] G. Kramer, I. Marić, and R. D. Yates, *Cooperative Communications*. Now Publishers Inc., 2006, vol. 1, no. 3.
- [6] M. Cardone, D. Tuninetti, R. Knopp, and U. Salim, "On the Gaussian Half-Duplex Relay Channel," *IEEE Trans. Inform. Theory*, vol. 60, pp. 2542–2562, May 2014.
- [7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 2012.
- [8] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [9] N. Zlatanov, V. Jamali, and R. Schober, "Capacity of the Two-Hop Half-Duplex Relay Channel," *Extended Version, Submitted to IEEE Trans. Inform. Theory*. [Online]. Available: <http://arxiv.org/abs/1411.5299>